

Ardenne S. Martin, L2
LSU Health Sciences Center, New Orleans, LA

Mentor's Name: Dr. Bradley Spieler
LSUHSC, Department of Radiology

“The Relevance of More Training Data on Accuracy of Model Prediction on Chest X-RAY”

Background: Artificial Intelligence (AI) has the potential to augment radiologist workflow and decrease strain on radiologists. Development of deep neural networks, artificial neural networks with neurons, synapses, weights, and functions, can be used to train AI models, serving as a response to this need. Publicly available databases like MIMIC-CXR and ChestX-ray8 offer hundreds of thousands of chest x-rays (CXR) used to train neural network models to detect and localize devices or disease. These publicly available databases make training AI models on large datasets possible. Once models are aptly trained, however, challenges include implementation into hospital Picture Archiving and Communication Systems (PACS), communication between physicians and information technology (IT) groups at hospitals, Health Insurance Portability and Accountability Act (HIPAA), privacy or medico-legal issues, translation into the workflow of experienced physicians, and funding. Ultimately, implementation of working models will allow radiologists to not only review more cases but to focus on more complex cases while maintaining normal workflow.

Purpose: To examine variations in dataset size and associated performance by convolutional neural networks (CNN) in detection and localization of devices or disease on CXR. We also introduce our approach to training a CNN regression model on CXRs of intubated patients from the publicly available MIMIC-CXR database to automatically detect endotracheal tube (ETT) position above the carina.

Materials and Methods: We conducted a retrospective review of radiology reports of CXRs of intubated patients using the publicly available MIMIC-CXR database. Our initial model used various numbers of training and validation images up to 7,000 images. These data were run for 50 epochs with a batch size of 16 and 300 x 300-pixel images with image augmentation including rotation (up to 30 degrees), perspective warping, magnification (up to 15% above baseline), and brightness and contrast adjustments (up to 25%). Our second model experimented with standardization of training with increasing dataset size, using a maximum of 5,000 chest radiograph images which were given 160,000 glances of the ETT position, with the number of epochs decreasing as the dataset increases. Finally, we examine other groups who studied binary output vectors with different prediction categories using both small and large datasets as well as those who studied multi-label outputs.

Results: Our first neural network regression model for endotracheal tube position was able to achieve mean absolute error (MAE) less than 1 cm on a MIMIC-CXR test dataset with area under the receiver operating curve of 0.911 for detecting tip position less than 2 cm above the carina. Our second model achieved similar results with MAE of 1.04 cm on a test dataset with area under the receiver operating curve of 0.922 for detecting tip position less than 2 cm above

the carina. After examination of 30 studies published since 2002, AUC was found to be highest for those that used 600-2,000 images as well as those that used 9,000-76,000 images.

Conclusion: Our results show that MAE decreased as our training dataset increased as evidenced by correlation with the test dataset in both the first and second models. Determining whether better outcomes are found in larger datasets will require further exploration and scrutiny of study design such as use of synthetic data, image augmentation techniques, localization of devices or disease, whether different databases were used for training and testing, whether networks were pre-trained or untrained, and disease and demographic characteristics of CXRs on which the models were trained.